Compte rendu TP RB

Hammami Firas

04/11/2014

I. Interface de Cytoscape

A.Control Panel

Il conteint les onglets suivants:

1. Network
2. VizMapper
3. Editor
4. Filters

Network

Il permet de lister les réseaux ayant été importés et de voir les caractéristiques en terme de nombre de noeuds et d'intéractions.

VizMapper

Cet onglet permet de personnaliser la représentation des noeuds et des arêtes (en choisissant leurs couleurs et formes).

Editor

Cet onglet permet d'éditer le réseau (en jouant sur l'ajout/délétion de noeuds et d'intéractions).

Filter

Il permet de filtrer les noeuds qu'on aura choisis selon plusieurs critères.

Data Panel

Il contient les onglets suivants:

1. Node Attribute Browser
2. Edge Attribute Browser
3. Network Attribute Browser

Node Attribute Browser

Il permet, à partir d'une sélection de noeuds, d'afficher leurs attributs (tels que l'ID[nom du noeud] ou le nom canonique).

Edge Attribute Browser

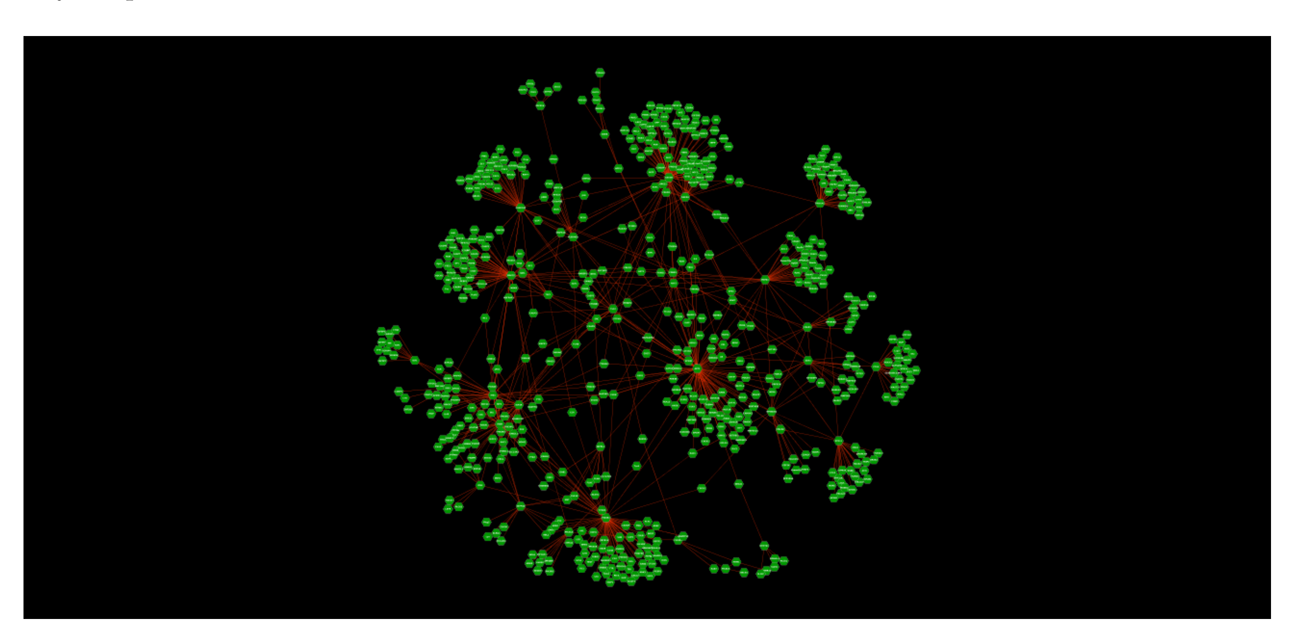
Il permet, à partir d'une sélection d'arêtes, d'afficher leurs atrtributs : $ID_{n1}(pp)ID_{n2}$

Avec $n1$ le noeud 1 et $n2$ le noeud 2.

Network Attribute Browser

Il permet d'afficher les attributs du réseau, tels que les noeuds parents ou fils [mais je n'ai pas vu de relation de parenté entre les noeuds].

Layout préféré



Tutorial pour colorer les arêtes découvertes par le crible Y2H ('This Study' = yes)

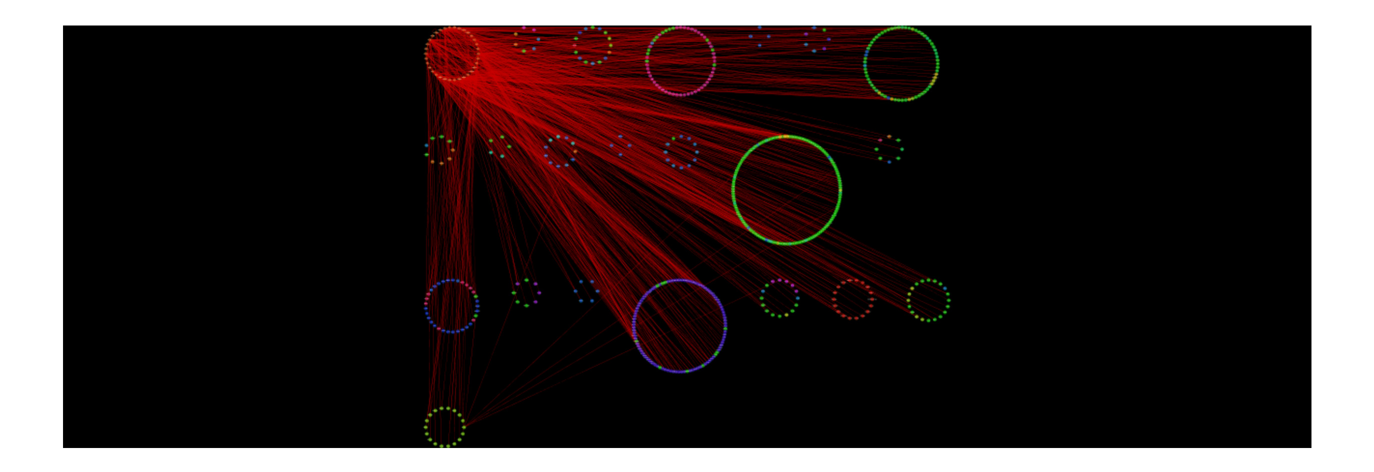
Dans l'onglet *VizMapper*,et plus particulièrement dans le panel *Visual Mapping Browser*:

1. Sélectionner *"Node Color"*
2. Mettre en paramètre de Mapping *"Discrete Mapping"*
3. Node Color : Selectionner *"Role in Y2H"*
4. Colorer avec deux couleurs différentes (Personnalisable) les *baits* et *preys*.

Visualisation du réseau Y2H

Paramètres utilisés:

1. Coloration *molecular function* (Rainbow)
2. Layout organisé selon les *subcellular functions*
3. Baits sous forme de *Diamond* (Présents en haut à gauche de la figure)



Le réseau fils contenant les noeuds ayant une fonction moléculaire inconnue

Ce réseau contient 34 noeuds, et 0 arêtes.

BINGO output

Le réseau obtenu représente les GeneOntology (GO) et leurs héritages, tout en colorant les GO les plus représentées en orangé ($p < 5 \cdot 10^{-7}$) et les GO les moins représentées en jaune.

Les 10 GO les plus représentées

Les 10 1res valeurs en terme de p-value:

```r
p <- vector()
p["response.to.insulin.stimulus"] <- 3.48e-16
p["phosphate.metabolic.process"] <- 3.48e-16
p["protein.kinase.B.signaling.cascade"] <- 6.15e-15
p["response.to.peptide.hormone.stimulus"] <- 2.28e-14
p["phosphorylation"] <- 6.98e-14
p["signaling"] <- 1.75e-13
p["regulation.of.glucose.import"] <- 1.75e-13
p["signaling.pathway"] <-6.53e-13
p["regulation.of.glucose.transport"] <-6.53e-13
p["protein.amino.acid.phosphorylation"] <-8.95e-13

print(as.data.frame(p))
```

```
##                                             p
## response.to.insulin.stimulus         3.48e-16
## phosphate.metabolic.process          3.48e-16
## protein.kinase.B.signaling.cascade   6.15e-15
## response.to.peptide.hormone.stimulus 2.28e-14
## phosphorylation                      6.98e-14
## signaling                            1.75e-13
## regulation.of.glucose.import         1.75e-13
## signaling.pathway                    6.53e-13
## regulation.of.glucose.transport      6.53e-13
## protein.amino.acid.phosphorylation   8.95e-13
```

Les 10 GO les plus sous-représentés

```r
p.val <- vector()
p.val["negative.regulation.of.JNK.cascade"] <- 4.93e-2
p.val["nitric.oxide.metabolic.process"] <- 4.93e-2
p.val["regulation.of.tyrosine.phosphorylation.of.Stat5.protein"] <- 4.93e-2
p.val["regulation.of.cholesterol.metabolic.process"] <- 4.93e-2
p.val["somatic.stem.cell.maintenance"]<-4.93e-2
p.val["positive.regulation.of.cyclin-dependent.protein.kinase.activity"] <-4.93e-2
p.val["negative.regulation.of.osteoclast.differenciation"] <- 4.93e-2
p.val["positive.regulation.of.erythrocyte.differenciation"] <-4.93e-2
p.val["regulation.of.myeloid.cell.apoptosis"] <- 4.93e-2
p.val["chondroitin.sulfate.proteoglycan.biosynthetic.process"] <-4.93e-2

print(as.data.frame(p.val))
```

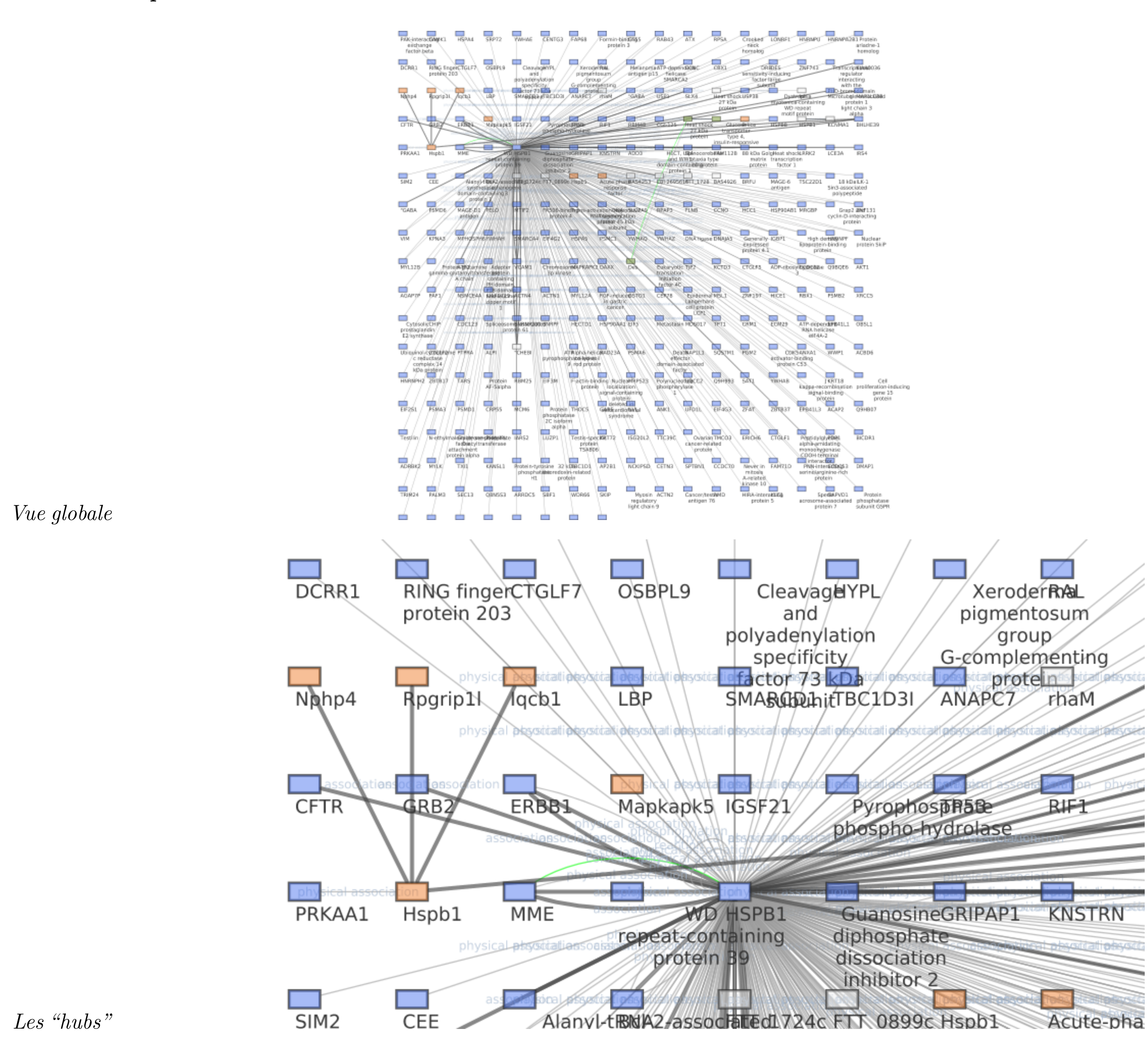
```
##                                                                   p.val
## negative.regulation.of.JNK.cascade                               0.0493
## nitric.oxide.metabolic.process                                   0.0493
## regulation.of.tyrosine.phosphorylation.of.Stat5.protein          0.0493
## regulation.of.cholesterol.metabolic.process                      0.0493
## somatic.stem.cell.maintenance                                    0.0493
## positive.regulation.of.cyclin-dependent.protein.kinase.activity 0.0493
## negative.regulation.of.osteoclast.differenciation                0.0493
## positive.regulation.of.erythrocyte.differenciation               0.0493
## regulation.of.myeloid.cell.apoptosis                             0.0493
## chondroitin.sulfate.proteoglycan.biosynthetic.process            0.0493
```

Interprétation

Globalement, on peut voir que les noeuds présentent un enrichissement en GO du type *insulin pathway*, *signalling*, ou encore de type *transport de glucose*. Ces résultats suggèrent donc que ces gènes sont impliqués dans la voie de signalisation de l'insuline.

II) Création d'un réseau d'intéraction pour une protéine d'intérêt (HSPB1)

Caractéristiques du réseau Intact



Vue globale

Les "hubs"

Ce réseau de 297 noeuds présente 359 arêtes.

Parmi ces noeuds, on retrouve deux hubs, qui sont **WD HSPB1 repeats-ontaining domain 39** et **HSPB1** (mais avec beaucoup moins d'arêtes).

On peut voir que le noeud *WD HSPB1 repeats-ontaining domain 39* présente une pléthore d'arêtes, qui relient des noeuds qui pour la plupart présentent un unique lien (celui avec le hub).

III) Le partitionnement du réseau

Analyse des partitions réalisées par le Clust&See

Cluster 3 [Whole annotation vs Network reference]

GO	C3 whole annotation	p-value corrected	C3 Network reference	p-value corrected
1rst	TM receptor protein tyrosine kinase signaling	$3,95 \cdot 10^{-25}$	TM receptor protein tyrosine kinase signaling	$1,49 \cdot 10^{-5}$
2nd	enzyme linked receptor protein signaling pathway	$3,02 \cdot 10^{-23}$	enzyme linked receptor protein signaling pathway	$1,68 \cdot 10^{-5}$
3rd	signaling	$7,1 \cdot 10^{-23}$	signaling	$4,33 \cdot 10^{-4}$
4th	signaling pathway	$2,84 \cdot 10^{-22}$	signaling pathway	$4,81 \cdot 10^{-4}$
5th	insulin receptor signaling pathway	$1,6 \cdot 10^{-19}$	cell surface receptor linked signaling pathway	$8,32 \cdot 10^{-4}$
6th	cellular response to peptide hormone stimulus	$9,06 \cdot 10^{-18}$	insulin receptor signaling pathway	$1,67 \cdot 10^{-3}$
7th	signal trandsduction	$7,94 \cdot 10^{-17}$	cellular response to peptide hormone stimulus	$2,93 \cdot 10^{-3}$
8th	response to hormone stimulus	$7,94 \cdot 10^{-17}$	response to hormone stimulus	$8,42 \cdot 10^{-3}$
9th	cell surface receptor linked sgnaling pathway	$9,53 \cdot 10^{-17}$	cellular response to endogenous stimulus	$9,21 \cdot 10^{-3}$
10th	response to insulin stimulus	$2,89 \cdot 10^{-16}$	response to peptide hormone stimulus	$9,21 \cdot 10^{-3}$

Déjà il semble qu'utiliser l'option *Whole Annotation* permet d'obtenir de faibles p-values corrigées, et les résultats utilisant l'autre option sont quasiment les mêmes dans ce cluster (sans compter la différence d'environ 10^{20} entre les p-values corrigées).

Cluster 4 [Whole annotation vs Network reference]

GO	C4 whole annotation	p-value corrected	C4 Network reference	p-value corrected
1rst	regulation of phosphorylation	$2,97 \cdot 10^{-5}$	small molecule catabolic process	$4,93 \cdot 10^{-2}$
2nd	regulation of phosphorus metabolic process	$2,97 \cdot 10^{-5}$	none	NA
3rd	regulation of phosphate metabolic process	$2,97 \cdot 10^{-5}$	none	NA
4th	anti-apoptosis	$2,97 \cdot 10^{-5}$	none	NA
5th	negative regulation of cellular process	$2,97 \cdot 10^{-5}$	none	NA
6th	regulation of apoptosis	$2,97 \cdot 10^{-5}$	none	NA

GO	C4 whole annotation	p-value corrected	C4 Network reference	p-value corrected
7th	negative regulation of apoptosis	$2,97 \cdot 10^{-5}$	none	NA
8th	regulation of programmed cell death	$2,97 \cdot 10^{-5}$	none	NA
9th	regulation of cell death	$2,97 \cdot 10^{-5}$	none	NA
10th	negative regulation of cell death	$2,97 \cdot 10^{-5}$	none	NA

Le cluster C4 indique de hautes p-values corrigées (de l'ordre de 10^{-5}) en utilisant l'option *Whole annotation*. De plus en changeant cette option, on se retrouve avec un seul résultat en 1er rang.

Cluster 10 [Whole annotation vs Network reference]

GO	C10 whole annotation	p-value corrected	C10 Network reference	p-value corrected
1rst	intracellular signaling pathway	$1,62 \cdot 10^{-3}$	none	NA
2nd	signaling pathway	$1,62 \cdot 10^{-3}$	none	NA
3rd	regulation of apoptosis	$2,02 \cdot 10^{-3}$	none	NA
4th	regulation of programmed cell death	$2,02 \cdot 10^{-3}$	none	NA
5th	regulation of cell death	$2,02 \cdot 10^{-3}$	none	NA
6th	negative regulation of biological process	$2,02 \cdot 10^{-3}$	none	NA
7th	response to dsRNA	$2,02 \cdot 10^{-3}$	none	NA
8th	regulation of cell communication	$2,1 \cdot 10^{-3}$	none	NA
9th	anti-apoptosis	$2,27 \cdot 10^{-3}$	none	NA
10th	phospate metabolic process	$2,27 \cdot 10^{-3}$	none	NA

Ce cluster indique qu'avec l'option *Network Reference*, BINGO n'a pas pu trouver de résultat. De plus, l'utilisation de *Whole Reference* indique des p-values de l'ordre de 10^{-3}. Ces valeurs de p-values obtenues ne peuvent donc pas être significatifs.
Cette hypothèse est vérifiable par la modification d'un paramètre de sélection des catégories à visualiser => *All Categories*. On obtient le résultat suivant (en utilisant l'option *Use Network as reference set*):

GO rank	C10 Network reference	Corrected p-value
1rst	response to dsRNA	$9.15 \cdot 10^{-2}$
2nd	positive regulation of cell morphogenesis involved in differentiation	$1,43 \cdot 10^{-1}$
3rd	production of miRNAs involved in gene silencing by miRNA	$1,43 \cdot 10^{-1}$
4th	gene silencing by miRNA	$1,43 \cdot 10^{-1}$
5th	posttranscriptional gene silencing by RNA	$1,43 \cdot 10^{-1}$
6th	regulation of epithelial to mesenchymal transition	$1,43 \cdot 10^{-1}$

GO rank	C10 Network reference	Corrected p-value
7th	embryonic cranial skeleton morphogenesis	$1,43 \cdot 10^{-1}$
8th	posttranscriptional gene silencing	$1,43 \cdot 10^{-1}$
9th	cellular response to dsRNA	$1,43 \cdot 10^{-1}$
10th	pericardium development	$1,43 \cdot 10^{-1}$

Comme on peut le voir, les GO sont diverses et variées et leurs p-value corrigées sont proches de 0,1 , signifiant qu'elles ont été trouvées au hasard. Lors de la visualisation des catégories surreprésentées après correction, ces résultats sont masqués.

Overlapping Cluster generator

La protéine Multiclassée

La protéine multiclassée (située entre les clusters C3 et C7) corresponds à GSK3B.

Comparaison des clusters C3 et C7

GO rank	C3 Annotation	Corrected p-value	GO rank	C7 Annotation	Corrected p-value
1rst	regulation of cellular process	$2,82 \cdot 10^{-7}$	1rst	histone H3-T6 phosphorylation	$1,7 \cdot 10^{-6}$
2nd	regulation of biological process	$8,63 \cdot 10^{-7}$	2nd	protein amino acid phosphorylation	$1,7 \cdot 10^{-6}$
3rd	regulation of apoptosis	$2,21 \cdot 10^{-6}$	3rd	negative regulation of insulin receptor signaling pathway	$1,7 \cdot 10^{-6}$
4th	regulation of programmed cell death	$2,21 \cdot 10^{-6}$	4th	phosphorylation	$1,7 \cdot 10^{-6}$
5th	biological regulation	$2,21 \cdot 10^{-6}$	5th	negative regulation of glucose import	$3,09 \cdot 10^{-6}$
6th	regulation of cell death	$2,21 \cdot 10^{-6}$	6th	histone phosphorylation	$5,01 \cdot 10^{-6}$
7th	signaling	$4,55 \cdot 10^{-6}$	7th	negative regulation of glucose transport	$8,5 \cdot 10^{-6}$
8th	regulation of developmental process	$4,55 \cdot 10^{-6}$	8th	phosphate metabolic process	$1,05 \cdot 10^{-5}$
9th	positive regulation of protein complex assembly	$4,73 \cdot 10^{-6}$	9th	phosphorus metabolic process	$1,05 \cdot 10^{-5}$
10th	multicellular organismal development	$6,3 \cdot 10^{-6}$	10th	regulation of response to stimulus	$1,6 \cdot 10^{-5}$

A priori les Gene Ontologies obtenues par l'annotation via BINGO sont très différentes entre C3 et C7. On peut donc émettre l'hypothèse que GSK3B soit un hub entre ces deux classes, et que ce noeud soit une interface entre deux types de processus biologiques différents (*régulation de processus développementaux* **vs** *régulationde la voie de signalisation de l'insuline*)